Spatial Modeling Of Counts

J. Andrew Royle
USGS Patuxent Wildlife Research Center

Workshop on Spatial Statistics
Beltsville Agricultural Research Center
March 16, 2006

Outline

- Introduction: Count data in ecology and spatial dependence
- Generalized Linear Modeling (GLM) framework
- Spatial correlation models
- Examples: North American BBS data
- Detection bias in animal surveys

Introduction

Ecology: *The study of spatial and temporal variation in abundance*

A general theme of ecological studies: Collect spatially referenced counts, $y(s)$, with the goal of making inferences about "abundance"

For example,

- Characterize the spatial distribution of a population
- Map occurrence of a species – "range map"
- Evaluate landscape factors that influence variation in abundance

Introduction

Data: $y(s_i) \equiv y_i$ are spatially referenced *counts*, e.g., number of birds counted at site s_i (a point, quadrat, transect)

Genesis of Spatial Dependence –

- Omitted habitat covariates
- Demographic processes

  $\rightarrow$ Recruitment, dispersal, etc..

- Interactions between individuals/species

  $\rightarrow$ Predation, competition

Objectives

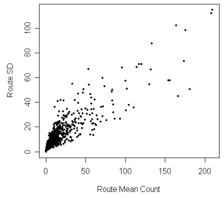
What do we do with spatial models of abundance?

- Mapping/prediction or simple description
- Small area estimation, inference
- Shrinkage estimation of model parameters
- "Honest" estimation of covariate effects

Considerations for Modeling Counts

Why not just use a kriging-type model?

- counts are positive valued
- counts are discrete
- mean related to variance (empirically)



← Route SD vs. mean, house finch (routes ≥ 10 years)

Kriging is a linear procedure, for normally distributed data that does not respect these features.

Generalized Linear Models (GLMs):

Classical statistics deals with normal distributions and linear models.

- $y_i \sim \text{Normal}(\mu_i, \sigma^2)$
- $\mu_i = \beta_0 + \beta_1 x_i$

Kriging is also a normal, linear procedure

GLMs (Generalized Linear Models) represent an analogous class of models for non-normal data

Elements of Generalized Linear Models (GLMs)

A probability model for the observations:

- $f(\mu_i, \theta)$
  - $\mu_i = E[y_i]$
  - θ = a variance parameter

Common choices of f for count data

- Poisson
- Binomial

Generalized Linear Models (GLMs)

Modeling covariates effects:

$$h(E[y_i]) = \sum_{j=1}^{J} \beta_j x_{ij}$$

instead of (for normal data)

$$E[y_i] = \sum_{j=1}^{J} \beta_j x_{ij}$$

- $h(\cdot)$ is called the *link* function (it *links* the mean of $f(\cdot)$ to the linear function of covariates)
  - Poisson: $log(\mu_i)$
  - Binomial: $log(\mu_i/(1-\mu_i))$

Poisson Regression

Probability model for the data:

$y_i \sim \text{Poisson}(\mu_i)$

μ_i is the mean of y_i at location s_i

$log(\mu_i) = \beta_0 + \beta_1 x_i$

x_i = a covariate, describing landscape or habitat structure

GLMs for Spatial Data

Introduce a spatially indexed random effect, z_i:

$$h(\mu_i) = \sum_{j=1}^{J} \beta_j x_{ij} + z_i$$

- z_i is a spatially correlated random effect
- Exploit conventional Gaussian spatial process models for z_i (kriging)
- Several possibilities are described shortly

Binomial counts

If y is the number of "successes" in T independent Bernoulli trials ("coin flips"), then y has a binomial distribution

- T = sample size
- parameter π = "success probability"

Binomial data examples

- Nest success/productivity data
- Capture-recapture or band recovery data
- Occupancy data (y_i units occupied out of T_i)
- Harvest success

Binomial counts

Goal: model variation in π_i

Logistic regression model:

$$log(\pi_i/(1-\pi_i)) = \sum_{j=1}^{J} \beta_j x_{ij} + z_i$$

Poisson Counts

Aggregate a Poisson point process (equal area units)

$$y_i \sim \text{Poisson}(\mu_i)$$

y_i results from counting (unique) individuals in space

Goal: model variation in μ_i

Log-linear model:

$$log(\mu_i) = \sum_{j=1}^{J} \beta_j x_{ij} + z_i$$

Spatial Models for z —

Assume that $z_i \equiv z(s_i)$ is a Gaussian spatial process:

- $z_i \sim$ Normal
- $E[z_i] = 0$
- $Var[z_i] = \sigma^2$
- $Corr(z_i, z_j) = k_\theta(||s_i - s_j||)$

Joint normality of $\mathbf{z} = (z_1, z_2, \ldots, z_n)$:

$$\mathbf{z}_{n\times 1} \sim \text{Normal}(0, \mathbf{\Sigma}(\theta))$$

There are a number of ways to specify $\mathbf{\Sigma}(\theta)$

1. Classical or Direct Construction

"Kriging for counts" – A direct specification of a joint distribution for the spatial process, $z(s)$

Specify a model for the correlation between $z(s)$ at any two locations:

$$Corr(z(s_i), z(s_j)) = k_\theta(||s_i - s_j||)$$

e.g., exponential decay –

$$k_\theta(s, s') = e^{-||s-s'||/\theta}$$

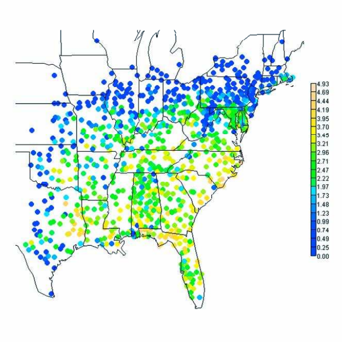
This function $k_\theta(s, s')$ "fills-in" the $n \times n$ elements of $\mathbf{\Sigma}(\theta)$:

$$\mathbf{z}_{n\times 1} \sim \text{Normal}(0, \mathbf{\Sigma}(\theta))$$

Estimation/prediction requires repeated mathematical operations on $\mathbf{\Sigma}(\theta)$

Example: Range Mapping

- Carolina Wren counts from the BBS
- abt. 1000 routes
- Goal is to make a relative abundance/range map



← log(mean) route count (across years)

$\mathbf{\Sigma}(\theta)$ is 1000×1000 and does not yield to kriging-like estimation and prediction.

Kriging for Counts

Diggle, P.J., J.A. Tawn and R.A. Moyeed. 1998. Model-based geostatistics. *Journal of the Royal Statistical Society, Ser. C.*

2. Kernel Smoothing/(Process Convolution) Construction

Express $z(s)$ as a linear combination of *iid* "random effects"

$$z(s) = \sum_{j=1}^{R} w_\theta(r, s)\alpha(r_j)$$

where

$$\alpha(r) \sim \text{Normal}(0, \sigma^2_\alpha)$$

- $w_\theta(r, \cdot)$ is a kernel centered at r
  "kernel" = weighting function

- z an average of "noise" –
  $z(s)$ is a weighted average of *iid* noise $\alpha(r_j); j = 1, 2, \ldots, R$.

- A classical mixed model (Laird and Ware; PROC MIXED)

- $R << n$

Kernel Smoothing/Convolution Construction

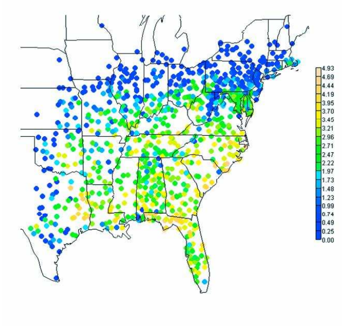
- Equivalence between this method and "kriging", i.e., a precise relationship between the choice of $w_\theta(\cdot)$ and the correlation function.

- This is more computationally efficient in large problems. Do not have to operate on $\mathbf{\Sigma}(\theta)_{n\times n}$.

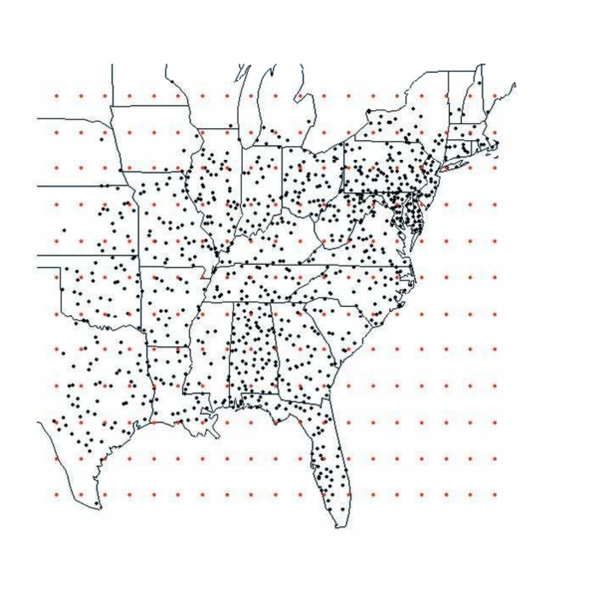
- Higon, D. 1998. A process-convolution approach to modeling temperatures in the North Atlantic Ocean. *Environmental and Ecological Statistics*

Example: Range Mapping

- Carolina Wren counts from the BBS
- abt. 1000 routes
- Goal is to make a relative abundance/range map
- Method: Gaussian kernel convolution model

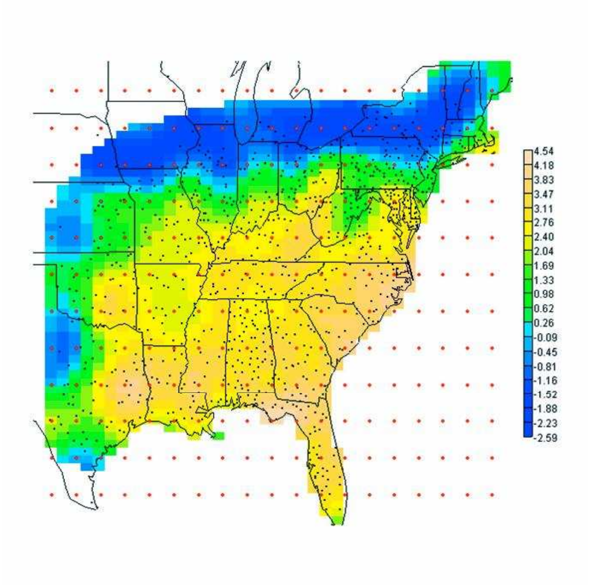


← log(mean) route count (across years)



← Data locations and grid of "support points" – Gaussian kernel centers

Estimated spatial process:



3. Lattice models

Usually used when *data* have discrete or areal support. e.g., areal measurements: counties, geographic strata, etc..

Conditional autoregression (CAR):

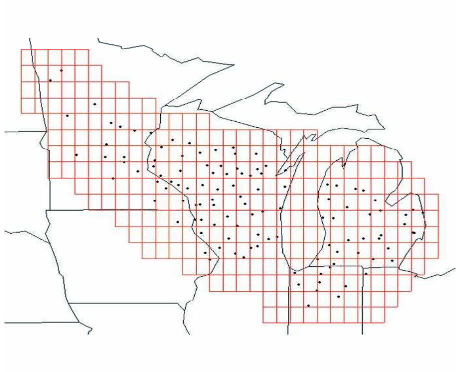
$$z_i = \rho \sum_{j \sim i} w_{ij} z_j + \epsilon_i$$

$\{w_{ij}\} \equiv \mathbf{W}$ is the *adjacency* matrix.

- 0s and 1s indicating neighbors
- length of boundary
- "average distance" between cells

Lattice models for non-lattice data

If data locations do not form a natural lattice, then make one up:



$log(\boldsymbol{\mu}) = \mu\mathbf{1} + \mathbf{Hz}$

- $\boldsymbol{\mu}$ is $n \times 1$
- $\mathbf{z}$ is $p \times 1$ CAR process
- $\mathbf{H}$ is $n \times p$

$\mathbf{H}$ associates each observation with one or more of the p random effects, which are arranged on a lattice

BBS Bobolink counts, arbitrary grid for embedded CAR model

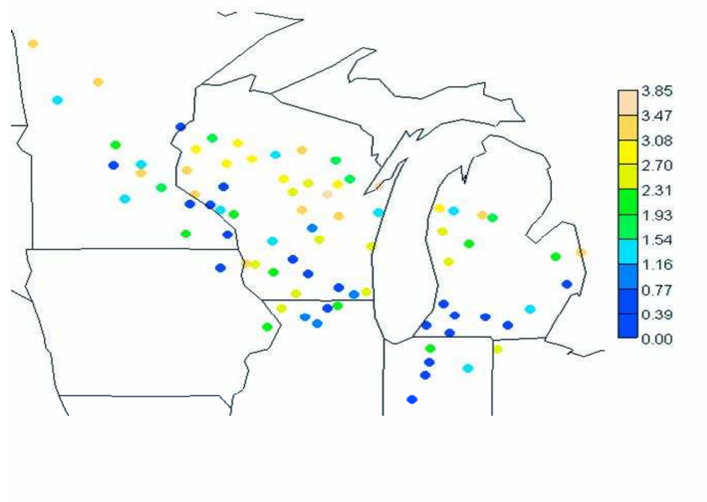
Example: Spatial Variation in Bobolink Counts

- Species: Bobolink
- BBS route counts in the upper-midwest (a physiographic stratum)
- Several habitat covariates thought to influence abundance
- CAR model with incidence adjacency matrix

Data Locations

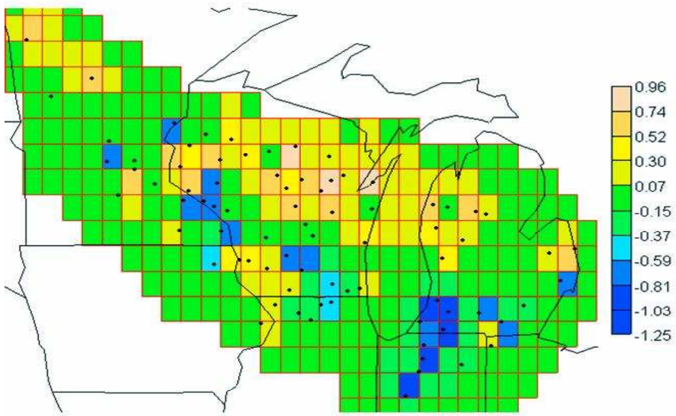
100 or so routes in upper midwest
y_i = count of bobolinks on BBS route i, located at s_i.

Data



$log(count)$

Predictions



Estimation and Implementation

- Markov chain Monte Carlo

geoR,geoRGLM add-on **R** libraries

PROC MIXED/GLIMMIX for some models

WinBUGS for all models described here

Abundance and Detectability

In Ecology, we have an acute inability to observe the state variable of interest in many problems: Abundance, or occurrence

$N(s) = \#$ of animals in population s (population size)

Observe a sample count, $y(s) \leq N(s)$

Abundance and Detectability

Binomial Observation Model:

$$y(s) \sim \text{Binomial}(N(s), p)$$

$$\begin{aligned} y(s) &= \text{observed count} \\ p &= \text{"detection probability"} \end{aligned}$$

- Detection is important because y is a "biased estimate" of N
- p can vary in response to many factors (e.g., intensity, env. conditions)
- Variation in y is not just due to variation in N.
- But (variation in) N is the object of inference

Simple Count Surveys (Binomial counts)

When detection is imperfect, $N(s)$ is not distinguishable from p (they are confounded). For example, the model consisting of:

(1) $y(s) \sim \text{Binomial}(N(s), p)$ and

(2) $N(s) \sim \text{Poisson}(\mu(s))$

is equivalent to the model

$$y(s) \sim \text{Poisson}(p\mu(s))$$

Thus, models for $y(s)$ describe variation in the product $p\mu(s)$. This is insufficient for some important inference problems.

Abundance and Detection

Therefore, much effort has been directed toward developing alternative sampling protocols/methods that allow variation due to the detection process to be decoupled from variation in abundance.

- capture-recapture
- double or multiple observer sampling
- distance sampling
- "removal" methods

Most methods yield a multivariate count statistic $\mathbf{y}$ that has a multinomial sampling distribution –

$$\mathbf{y}|N \sim \text{Multinomial}(N; \boldsymbol{\pi})$$

Differences among protocols are manifest in parameterization of $\boldsymbol{\pi}$

Example of Multinomial Observation Models

A double-observer protocol: Two observers independent record observations of individuals and, after the fact, "reconcile" their observation lists. This yields an *encounter history* for each individual of the form:

1	1	observed by both observers
1	0	observed by 1st
0	1	observed by 2nd
0	0	not observed

Data are encounter history *frequencies* – n_{11}, n_{10}, n_{01} and n_{00} (missing data), which have a multinomial distribution, with cell probabilities π_{11}, π_{10}, π_{01}, π_{00}. These are functions of detection probability p_1 (1st observer) and p_2 (2nd observer).

The General Hierarchical Model

1. Multinomial Likelihood –

$$\mathbf{y}|N \sim \text{Multinomial}(N; \boldsymbol{\pi})$$

2. Abundance model –

$$N_i \sim Poisson(\mu_i)$$

3. Model for the Poisson mean

$$log(\mu_i) = \mathbf{x}_i'\boldsymbol{b} + z(s_i)$$

4. The spatial process – Spatial dependence is induced through the correlated random effect, $z(s)$.

Summary

- Many ecological studies yield data that are counts: of animals, or Bernoulli trials
- Poisson/Binomial GLMs with spatially correlated random effects
  1. Kriging-type models
  2. Regression-on-noise ("convolution") formulation
  3. Lattice models (CAR)
- Abundance/occurrence processes, detection bias: yields a hierarchical model wherein the spatial model governs the latent (unobservable) abundance parameter, $N(s)$.